The additive generalization of the Boltzmann entropy

Alexander N. Gorban*

ETH Zürich, Department of Materials, Institute of Polymers,

ETH-Zentrum, CH-8092 Zürich, Switzerland

Institute of Computational Modeling RAS, 660036 Krasnoyarsk, Russia

Iliya V. Karlin† and Hans Christian Öttinger‡

ETH Zürich, Department of Materials, Institute of Polymers,

ETH-Zentrum, CH-8092 Zürich, Switzerland

(Dated: February 1, 2008)

Abstract

There exists only one generalization of the classical Boltzmann-Gibbs-Shannon entropy functional to a one-parametric family of additive entropy functionals. We find analytical solution to the corresponding extension of the classical ensembles, and discuss in some detail the example of the deformation of the uncorrelated state.

PACS numbers: 05.70.Ln, 05.20.Dd

*Electronic address: gorban@icm.krasn.ru

†Corresponding author; Electronic address: ikarlin@ifp.mat.ethz.ch

‡Electronic address: hco@ifp.mat.ethz.ch

I. INTRODUCTION

The growing interest to non-classical entropies in recent years [1, 2] is motivated by the fact that they can be used to describe observable statistical effects such as: (i) Non-classical tails of distribution functions which can deviate significantly from Gaussian distribution. In particular, this asymptotics can be power-law ("long tails") or, instead, distribution functions can decay in a more rapid fashion ("short tails"), in particular, they can become equal to zero at finite distance ("cut tails"). (ii). Strong correlations between subsystems in equilibrium and conditionally-equilibrium (quasi-equilibrium) states. (iii) In particular, even at a vanishingly weak interaction between subsystems, when the classical Boltzmann-Gibbs-Shannon entropy (BGS) would lead to no correlations, strong correlations can be observed in the equilibrium. This may sound somewhat paradoxal: Joining non-interacting subsystems with equal values of the intensive quantities, and switching on an infinitesimal weak interaction, we produce a strongly correlated equilibrium. However, the simplest example is readily provided (though not related to non-classical entropies *per se*) by the microcanonical ensemble of finite systems: If subsystems are not interacting at all, then there is an additional conservation law, the energies of the individual subsystem, and the product of the microcanonic distributions is the equilibrium. However, an arbitrarily weak interaction will surely destroy this conservation law, and the equilibrium becomes the usual microcanonic ensemble (the equipartition over the surface of constant total energy). For finite number of particles in the subsystems, this latter state is correlated, and it does *not* factor into the product of the microcanonical distributions of the subsystems. It is only in the thermodynamic limit where the theorem about the equivalence of the ensembles [3] states the tendency to zero of correlations of (almost) noninteracting subsystems (in the domain of its applicability, of course). We should remark that empirically found asymptotics of the distribution functions should be always treated with care since they can turn out to be "intermediate asymptotics" rather than true limits.

The entropic description of all these effects in the spirit of Gibbs ensembles is technically advantageous (same as any variational principle) but this is by far not the only merit. If the entropy is consistent with the kinetics, and varies monotonically in time, then a very useful construction becomes available. This is the conditional equilibrium (or quasi-equilibrium, with local equilibrium as a specific example). The quasi-equilibrium describes partially re-

laxed systems, according to the idea of the fast-slow decomposition of motions: Fast variables have almost reached equilibrium at almost fixed values of slow variables. Conditional equilibrium is described as the probability distribution which brings to maximum the entropy $S(p)$ at fixed values of the slow variables, $M = m(p)$:

$$S(p) \to \max,\ m(p) = M. \tag{1}$$

Usually, when one attempts to introduce non-classical entropies in order to use these advantages, there is a price to be paid. Non-classical entropies at use in most of the contemporary studies violate at least one of the following important and familiar properties of the BGS entropy: (i) Additivity: The entropy of the system which is composed of independent subsystems equals the sum of the entropies of the subsystems. (ii) Trace-form: The entropy is a sum over the states (see below). (iii) Concavity of the entropy. For example, the Tsallis entropy [1] is not additive, the Rényi entropy [4] is not of the trace form.

Violation of additivity cannot be motivated by the fact that "in reality, all subsystems are interacting" [5]. Indeed, the additivity axiom is the conditional statement: *If* the systems are independent, *then* the entropy of the joint system equals the sum of the entropies of subsystems. Probability theory, even when studying whatever strongly dependent events, is based on such notions as independence, independent trials etc [6]. Giving up these notions simply on the grounds that events in nature depend on each other is misleading.

In this paper we demonstrate how the description of both long and short tail distributions, growth of correlations etc can be achieved on the basis of the entropy approach, and without a violation of neither the additivity nor of the trace form requirements (however, with a violation of the concavity only for the description of cut tail distributions). Such a description becomes available only if one uses a one-parametric family of entropies introduced recently [7]. We establish analytic formulae for conditional maximizers of these entropies which makes operations within the present formalism almost as easy as in the case of a Gaussian distribution pertinent to the BGS entropy.

II. ADDITIVE TRACE-FORM ENTROPIES FOR MARKOV PROCESSES

The basic model we consider here is the finite Markov chain (finiteness and discreteness are by no means the crucial restriction, and are employed only in order to avoid the convergence

questions). The time evolution of the probabilities p_i, where i is the discrete label of the state, is given by master equation,

$$\dot{p}_i = \sum_{j,j\neq i} k_{ij}\left(\frac{p_j}{p_j^*} - \frac{p_i}{p_i^*}\right),\ k_{ij} = k_{ji} \geq 0. \tag{2}$$

We consider only systems which allow for a positive equilibrium, $p_i^* > 0$ (for infinite systems, it is often advantageous to use unnormalized p^*). We recall [7, 8] that, for each convex function of one variable, $h(x)$, one constructs the Lyapunov function $H_h(p)$ which does not increase on solutions to Eq. (2), where

$$H_h(p) = \sum_i p_i^* h(p_i/p_i^*). \tag{3}$$

[We consider below H_h-functions rather than entropy functions $S_h = -H_h$.]

Among the set of Lyapunov functions (3), there exists a one-parametric subset of additive Lyapunov functions, H_α, $0 \geq \alpha \geq 1$:

$$\begin{aligned} H_\alpha &= \sum_i p_i^* h_\alpha(p_i/p_i^*), \\ h_\alpha(x) &= (1-\alpha)x\ln x - \alpha \ln x. \end{aligned} \tag{4}$$

In particular,

$$\begin{aligned} H_0 &= \sum_i p_i \ln(p_i/p_i^*), \\ H_1 &= -\sum_i p_i^* \ln(p_i/p_i^*). \end{aligned}$$

Additivity of functions H_α (4) is readily checked [7, 8]: If $p = p_{ij} = q_i r_j$, and also if $p^* = p_{ij}^* = q_i^* r_j^*$, then

$$H_\alpha(p) = H_\alpha(q) + H_\alpha(r).$$

It can be demonstrated that the family (4) is unique (up to a constant factor): There are no other additive trace-form functions among Lyapunov functions (3) of master equation.

III. SOLUTION TO THE MAXIMUM ENTROPY PROBLEM

Since a factor in front of H_α is irrelevant, it proves convenient to use a different parameterization of the family (4),

$$H_\alpha = \sum_i [p_i \ln(p_i/p_i^*) - \alpha p_i^* \ln(p_i/p_i^*)], \tag{5}$$

where $\alpha \geq 0$, and where the case $\alpha \to \infty$ should be considered separately,

$$H_\infty = -\sum_i p_i^* \ln(p/p_i^*). \tag{6}$$

In order to address the construction of the quasi-equilibrium in a general setting, we assume the macroscopic variables $M = m(p)$, where $M_s = \sum_i m_{si} p_i$, and consider the problem (1) with $S = -H_\alpha$. Solving this problem with the method of Lagrange multipliers, we find:

$$\frac{\partial H_\alpha}{\partial p_i} = \lambda_0 + \sum_s \lambda_s m_{si}, \tag{7}$$

where Lagrange multiplier λ_0 corresponds to normalization, and λ_s to the rest of the constraints. Let us denote $-\Lambda_i$ the right hand side of Eq. (7). With this, Eq. (7) may be written,

$$\ln(p_i/p_i^*) - \alpha(p_i^*/p_i) = -\Lambda_i. \tag{8}$$

Solution to an equation,

$$\ln q - \alpha q^{-1} = -\Lambda, \tag{9}$$

may be written as follows:

$$q = e^{-\Lambda} e^{\mathrm{lm}\left(\alpha e^{\Lambda}\right)}, \tag{10}$$

where we have introduced notation lm$a$ (modified logarithm) for the function which is the solution to the transcendent equation,

$$x e^x = a.$$

The function lm satisfies the following identities:

$$\mathrm{lm}a \;=\; \ln a - \ln \mathrm{lm}a, \tag{11}$$

$$\mathrm{lm}a \;=\; \ln a - \ln(\ln a - \ln(\ln a - \ln(\dots))\dots). \tag{12}$$

Identity (12) is the recurrent application of identity (11). A different representation of solution (10) reads:

$$q = \frac{\alpha}{\mathrm{lm}(\alpha e^{\Lambda})}. \tag{13}$$

From the representation (10), the asymptotics at $\alpha \to 0$, and fixed Λ, is obvious: $q \to e^{-\Lambda}$, and which corresponds to the usual Boltzmann distribution. On the other hand, representation (13) reveals the asymptotics at $\Lambda \to \infty$:

$$q \sim \frac{\alpha}{\ln \alpha + \Lambda}.$$

For a symmetric distribution on the axis, and for $\Lambda = \lambda_0 + \lambda_2 x^2$, the first of the limits just mentioned gives the Gaussian distribution, while the second limit gives the Cauchy distribution. The corresponding distribution function for the limiting case H_∞ is simply the Cauchy distribution on the axis. Among non-symmetric Cauchy distributions of the form, $p = (\lambda_0 + \lambda_1 x + \lambda_2 x^2)^{-1}$, there are distinguished cases with a twice degenerated zero in the denominator: $p = (\lambda(x-a))^{-2}$. When one attempts to normalize this distribution by choosing a convergent sequence of functions, one gets a Dirac δ-function, $\delta(x-a)$ which can be interpreted as a microcanonic ensemble.

Thus, the quasi-equilibrium distribution has the form:

$$p = p^* e^{-\Lambda} e^{\mathrm{lm}(\alpha e^{\Lambda})} = \frac{\alpha p^*}{\mathrm{lm}(\alpha e^{\Lambda})}. \tag{14}$$

[We have omitted indices of states in p, p^*, and Λ.] Formula (14) is the main result of this paper.

It is also interesting to address the formal extension of the result (14) to negative α. Function $\mathrm{lm}a$ is defined and is continuous for $a \geq -e^{-1}$ ($\mathrm{lm}a \geq -1$). At $a \to -e^{-1}$, we have the limit, $d\mathrm{lm}a/da \to \infty$. If we formally extend, $\mathrm{lm}a = -\infty$ for $a < -e^{-1}$, then Eq (14) is a distribution with "cut tail". With this, there will be defined a non-zero ratio p/p^*:

$$\inf\{p/p^* | p \neq 0\} \geq |\alpha| > 0, \tag{15}$$

that is, either $p \geq |\alpha| p^*$, or $p = 0$. This construction is similar to a Maxwell construction of a stretched spinodal (the cut at the inflection point), and not to the global maximum of the entropy. Whereas such constructions are always necessary when working with non-convex thermodynamic potentials, will not further discuss the case $\alpha < 0$ in this paper.

IV. QUASI-EQUILIBRIUM ENSEMBLES NEAR THE BGS LIMIT

For the classical BGS entropy ($\alpha = 0$), the quasi-equilibrium distribution has the form:

$$p = p^* e^{-\Lambda}, \tag{16}$$

where Λ is the corresponding gradient of the entropy at the quasi-equilibrium, expressed in terms of Lagrange multipliers. Let us study the quasi-equilibrium (14) for small α. To the first order, we get:

$$p = p^* \left(e^{-\Lambda} + \alpha\right) + o(\alpha). \tag{17}$$

[Note that, in this expansion, dependence of Λ on the values of the macroscopic variables M is implicit. Explicit evaluation of this dependence requires, in addition, an expansion of Λ in terms of α which is used below when studying concrete examples.]

Substituting equation (17) into the kinetic equation (Markov chain in the present context, into the Liouville equation in the context of particle's dynamics, or, generally speaking, into the linear equation of the microscopic Markovian process), we easily see that the term αp^* gives no contribution to the resulting quasi-equilibrium dynamics. Indeed, we first notice that the relation between the time derivative $\dot{p}$ with the Lagrange multipliers is the same, as for the classical Boltzmann's distribution (16): If L is the linear operator of Markovian dynamics, $\dot{p} = Lp$, then, substituting for p on the right hand side of this equation the expression (17), and using linearity, we get

$$L\left(p^*\left(e^{-\Lambda}+\alpha\right)\right) = L\left(p^* e^{-\Lambda}\right).$$

Furthermore, defining the shifted macroscopic variables, $M_\alpha = m(p - \alpha p^*)$, we find that for the classical quasi-equilibrium dynamic equation, $dM/dt = m(Lp(M))$, where $Lp(M)$ is the microscopic vector field evaluated at the classical quasi-equilibrium states $p(M) = p^* \exp(-\Lambda(M))$ is affected only by a shift $M \to M_\alpha$, to the first order in α. In other words, the quasi-equilibrium dynamics of the ensemble (14) is driven by the classical dynamics resulting from the BGS entropy and Boltzmann distributions (16) to the first order in α.

In order to compute the quasi-equilibrium to second order in α, we must use the expansion of lm$a$ to third order,

$$\mathrm{lm}a = a - a^2 + (3/2)a^3 + o(a^3).$$

Then

$$p = p^*\left(e^{-\Lambda} + \alpha - \frac{1}{2}\alpha^2 e^{\Lambda}\right) + o(\alpha^2). \tag{18}$$

Further corrections can also be easily computed using higher-order terms in the expansion of the lm. We now shall consider a specific example of the formula (18).

V. EXAMPLE: ENHANCEMENT OF PARTICLE'S CORRELATIONS

In order to illustrate the effect of second order deviations from the BGS case, we apply Eq. (18) to the classical quasi-equilibrium defined by the one-particle configurational distribution function $f_1(r)$, where r is position variable. Assuming, as usual, the equipartition for the

reference equilibrium, $p^* = 1/V^N$, where V is the volume of the system, and N is the number of particles, we get $e^{-\Lambda} = e^{\lambda_0} \prod_{i=1}^{N} \Psi(r_i)$, where Lagrange multiplier λ_0 is responsible for normalization. Then the N-body quasi-equilibrium distribution function to second order in α reads,

$$V^N p = e^{\lambda_0} \prod_{i=1}^{N} \Psi(r_i) + \alpha - \frac{\alpha^2}{2e^{\lambda_0} \prod_{i=1}^{N} \Psi(r_i)} + o(\alpha^2). \tag{19}$$

Our goal now is to compute the two-body configurational distribution function,

$$f_2(r, q) = N(N-1) \int p(r, q, r_3, \ldots, r_N) dr_3 \ldots dr_N,$$

in the quasi-equilibrium (19). We recall that the classical result for the BGS entropy gives the uncorrelated two-body distribution, $f_2(r, q) \sim f_1(r) f_1(q)$, which also corresponds to the limit ($\alpha = 0$) of Eq. (19). Computation to the order α^2 requires expansion of Lagrange multipliers λ_0 and Ψ to the corresponding order. This computation is straightforward although tedious, thus we give here only the final result: The two-body quasi-equilibrium configurational distribution function f_2 reads:

$$\frac{N}{N-1} f_2(r, q) = (1 + \alpha + \alpha^2) \tilde{f}_1(r) \tilde{f}_1(q) + \alpha n^2 - \underline{\frac{\alpha^2}{2} n^2 B^N \varphi_1(r) \varphi_1(q)} + o(\alpha^2), \tag{20}$$

where $n = N/V$ is the average number density, and where we have introduced notation,

$$\tilde{f}_1(r) = f_1(r) - \alpha n, \tag{21}$$

$$\varphi_1(r) = \frac{f_1(r)}{n} - \frac{n}{B f_1(r)}, \tag{22}$$

$$B = \frac{1}{V} \int_V \frac{n}{f_1(r)} dr. \tag{23}$$

It is readily checked that the result (20) gives $f_2 = (N-1) N^{-1} f_1 f_1$ at $\alpha = 0$ which is identical with the classical uncorrelated pair distribution with correct normalization [9].

The first two terms in Eq. (20) amount again to the uncorrelated state with *homogeneously shifted* one-particle distributions ($\tilde{f}_1$ (21) instead of f_1, which amounts to a homogeneous subtraction of the average density times α). The underlined term (of the order of α^2), is the contribution responsible for correlations due to the use of the non-classical entropy. It also has a form of a product, but not of the distribution functions, rather, of functions of one variable (22). In order to see the effect of this term more explicitly, we assume

$$f_1(r) = n(1 + \zeta(r) N^{-1/2}), \tag{24}$$

where ζ is a function with zero average, and finite amplitude, $\langle\zeta\rangle = 0$, $\langle\zeta^2\rangle = \sigma^2$, where we have introduced notation for averaging over the volume, $\langle h\rangle = V^{-1}\int_V h dr$. Assuming large (but finite) number of particles, we find to the leading order in N:

$$B = 1 + \sigma^2 N^{-1} + o(N^{-1}),\ B^N = e^{\sigma^2} + o(1).$$

Thus,

$$\frac{N}{N-1} f_2(r,q) \approx (1+\alpha+\alpha^2)\tilde{f}_1(r)\tilde{f}_1(q) + \alpha n^2 - 2\alpha^2 n^2 \sigma^2 e^{\sigma^2} N^{-1}\theta(r)\theta(q), \qquad (25)$$

where we have denoted $\theta = \sigma^{-1}\zeta$, $\langle\theta^2\rangle = 1$. This correlation is negative once the sign of the deviations from the homogeneity at points r and q are the same, and positive if these deviations have the opposite signs.

VI. CONCLUSION

Once a classical statistical system is out of the thermodynamic limit, the exclusive character of the Boltzmann-Gibbs-Shannon entropy is fading away, and classical ensembles are not equivalent anymore. Whereas using the microcanonical ensemble for any description of finite systems may be most appropriate, this route is very complicated, at least from a computational standpoint. For that reason, seeking an entropic description of effects of finiteness is a relevant option.

We stress it once again, that the one-parametric family H_α, Eq. (4) and (5), is the *unique* generalization of the classical Boltzmann-Gibbs-Shannon entropy consistent with the additivity and the trace-form requirements simultaneously. It is reasonable therefore to study its applicability to a description of statistical systems out of the strict thermodynamic limit. The main result of this paper is the analytical description of the quasi-equilibria for this family of the entropy functions. We have demonstrated that the solutions to the entropy maximization problems are accessible in a fairly simple way, and which amounts to studying a function of one variable, $\mathrm{lm}a$. This makes studies of the non-classical ensembles described herein relatively uncomplicated, especially in the vicinity of the classical BGS solutions, where we expect, in the first place, the theory to be meaningful. Eventually, predictions can be compared in molecular dynamics simulations by making the size of the system smaller,

and/or the number of particles smaller. This is left for a future work.

---

[1] C. Tsallis, J. Stat. Phys. **52**, 479 (1988).

[2] http://tsallis.cat.cbpf.br/biblio.html.

[3] D. Ruelle, *Statistical Mechanics. Rigorous Results* (Benjamin, New York, 1969).

[4] A. Rényi, *Probability Theory* (North-Holland, Amsterdam, 1970).

[5] C. Beck, Europhys. Lett. **57**, 329 (2002).

[6] M. Kac, *Probability and Related Topics in Physical Sciences* (Interscience, NY, 1957).

[7] A. N. Gorban and I. V. Karlin, cond-mat/0205511.

[8] A. N. Gorban, *Equilibrium Encircling. Equations of Chemical Kinetics and Their Thermodynamic Analysis* (Nauka, Novosibirsk, 1984).

[9] N. N. Bugaenko, A. N. Gorban, and I. V. Karlin, Teor. Mat. Fiz. **88**, 430 (1991).